I CLAIM:

1. A method for measuring a strength of co-occurrence data, comprising:

extracting two or more chemical or biological molecules names from a database record from an inference database, wherein the inference database includes a plurality of inference database records created from an indexed literature database, and wherein the two or more chemical or biological molecule names co-occur in one or more records in an indexed scientific literature database;

determining a Likelihood statistic for a co-occurrence reflecting physico-chemical interactions between a first chemical or biological molecule name-A and a second chemical or biological molecule name-B extracted from the database record;

applying the Likelihood statistic to the co-occurrence to determine if the co-occurrence between the first chemical or biological molecule-A and the second chemical or biological molecule-B is a non-trivial co-occurrence reflecting physico-chemical interactions.

2. The method of Claim 1 further comprising a computer readable medium having stored therein instructions for causing a processor to execute the steps of method.

3. The method of Claim 1 wherein the step of determining a Likelihood statistic for a co-occurrence includes determining:

$$L_{AB} = P(A \mid B) * P(\neg A \mid \neg B) * P(B \mid A) * P(\neg B \mid \neg A),$$

wherein A and B are two chemical or biological molecule names which co-occur in one or more database records, wherein $P(A \mid B) \equiv$ (the probability of A given B), $P(B \mid A) \equiv$

(the probability of B given A), wherein $P(\neg A \mid \neg B) \equiv$ (the probability of not A given not B) and $P(\neg B \mid \neg A) \equiv$ (the probability of not B given not A).

4. The method of Claim 3 wherein $P(A \mid B)$ includes determining $c(AB) / c(B)$, wherein $c(AB) \equiv$ a number of database records in which A and B co-occur, and $c(B) \equiv$ a number of database records in which B occurs either with or without A.

5. The method of claim 3 wherein $P(B \mid A)$ includes determining $C(BA) / c(A)$, wherein $c(AB) \equiv$ a number of database records in which A and B co-occur and $c(A) \equiv$ a number of database records in which A occurs either with or without B.

6. The method of Claim 3 wherein $P(\neg A \mid \neg B)$ includes determining $(N - (c(A) + c(B) - c(AB))) / (N - c(B))$, wherein $N \equiv$ a total number of database records including co-occurrences of any chemical or biological molecule names, wherein $c(AB) \equiv$ a number of database records in which A and B co-occur, wherein $c(A) \equiv$ a number of database records in which A occurs either with or without B, and wherein $c(B) \equiv$ a number of database records in which B occurs either with or without A.

7. The method of Claim 1 wherein the step of applying the Likelihood statistic to determine if the co-occurrence between the first chemical or biological molecule-A and the second chemical or biological molecule-B is a non-trivial co-occurrence reflecting physico-chemical interactions includes applying the Likelihood statistic as a fractional

value between zero and one, wherein a value near zero indicates a trivial co-occurrence and a value near one indicates a non-trivial co-occurrence.

8. The method of Claim 1 wherein the step of applying the Likelihood statistic to determine if the co-occurrence between the first chemical or biological molecule-A and the second chemical or biological molecule-B is a non-trivial co-occurrence reflecting physico-chemical interactions includes applying the Likelihood statistic to determine if the co-occurrence between the first chemical or biological molecule-A and the second chemical or biological molecule-B is a non-trivial co-occurrence reflecting physico-chemical interactions in a cell.

9. A method for contextual querying of co-occurrence data, comprising:

selecting a target node from a first list of nodes connected by a plurality of arcs in a connection network, wherein the connection network includes a plurality of nodes representing a plurality of chemical or biological molecules names and a plurality of arcs connecting the plurality of nodes in a pre-determined order, and wherein the plurality of arcs represent co-occurrence values of physico-chemical interactions between chemical or biological molecules;

creating a second list of nodes by considering simultaneously a plurality of other nodes that are neighbors of the target node as well as neighbors of the plurality of other nodes in prior to the target node in the connection network;

selecting a next node from the second list of nodes using the co-occurrence values, wherein the next node is a most likely next node after the target node in the pre-determined order for the connection network based on the co-occurrence values.

10. The method of Claim 9 further comprising a computer readable medium having stored therein instructions for causing a processor to execute the steps of the method.

11. The method of Claim 9 wherein the plurality of arcs connecting the plurality of nodes in a pre-determined order includes a directed graph for a biological pathway.

12. The method of Claim 9 wherein the step of selecting a next node from the second list of nodes using the co-occurrence values includes selecting a next node in a biological pathway.

13. The method of Claim 9 wherein the co-occurrence values include Likelihood statistics.

14. The method of Claim 13 wherein the Likelihood statistics include Likelihood statistics calculated with:

$$L_{AB} = P(A \mid B) * P(\neg A \mid \neg B) * P(B \mid A) * P(\neg B \mid \neg A),$$

wherein A and B are two chemical or biological molecule names which co-occur in one or more database records, wherein $P(A \mid B) \equiv$ (the probability of A given B), $P(B \mid A) \equiv$

(the probability of B given A), wherein $P(\neg A \mid \neg B) \equiv$ (the probability of not A given not B) and $P(\neg B \mid \neg A) \equiv$ (the probability of not B given not A).

15. The method of Claim 9 wherein the co-occurrence values of physico-chemical interactions between chemical or biological molecules includes co-occurrence values of physico-chemical interactions between chemical or biological molecules in cells.

16. A method for query polling of co-occurrence data, comprising:

selecting a position in a connection network for an unknown target node from a first list of nodes, wherein the connection network includes a plurality of nodes representing a plurality of chemical or biological molecules names and a plurality of arcs connecting the plurality of nodes in a pre-determined order, and wherein the plurality of arcs represent co-occurrence values of physico-chemical interactions between chemical or biological molecules;

determining a second list of nodes prior to the position of unknown target node in the connection network;

determining a third list of nodes subsequent to the position of unknown target node in the connection network;

determining a fourth list of nodes included in both the second list of nodes and the third list of nodes; and

determining an identity for the unknown target node by selecting a node with a from the fourth list of nodes using a Likelihood statistic, wherein the Likelihood statistic includes a co-occurrence value reflecting physico-chemical interactions between a first chemical or biological molecule-A and a second chemical or biological molecule-B.

17. The method of Claim 16 further comprising a computer readable medium having stored therein instructions for causing a processor to execute the steps of the method.

18. The method of Claim 16 wherein the step of determining an identity for the unknown target node by selecting a node with a Likelihood statistic includes determining a Likelihood statistic with:

$$L_{AB} = P(A \mid B) * P(\neg A \mid \neg B) * P(B \mid A) * P(\neg B \mid \neg A),$$

wherein A and B are two chemical or biological molecule names which co-occur in one or more database records, wherein $P(A \mid B) \equiv$ (the probability of A given B), $P(B \mid A) \equiv$ (the probability of B given A), wherein $P(\neg A \mid \neg B) \equiv$ (the probability of not A given not B) and $P(\neg B \mid \neg A) \equiv$ (the probability of not B given not A).

19. The method of Claim 16 wherein the step of determining an identity for the unknown target node by selecting a node with a Likelihood statistic includes determining a simultaneous Likelihood statistic by selecting nodes in the fourth list of nodes, and for nodes from the fourth set of nodes, multiplying Likelihood statistics from the second set list of nodes by Likelihood statistics from the third list of nodes, and choosing a single node with the largest Likelihood statistic product value.

20. The method of Claim 16 wherein the step of determining an identity for the unknown target node by selecting a node with a Likelihood statistic includes determining a simultaneous Likelihood statistic by selecting nodes in the fourth list of nodes, and for nodes from the fourth set of nodes, adding Likelihood statistics from the second set list of nodes with Likelihood statistics from the third list of nodes, and choosing a single node with the largest Likelihood statistic summation value.

21. A method for creating automated biological inferences, comprising:

constructing a connection network using one or more database records from an inference database, wherein the connection network includes a plurality of nodes for chemical or biological molecules and biological processes found to co-occur one or more times, wherein the plurality of nodes are connected by a plurality of arcs in a pre-determined order, and wherein the inference database was created from chemical or biological molecule and biological process information extracted from a structured literature database;

applying Likelihood statistic analysis methods to the connection network to determine possible inferences between the chemical or biological molecules and biological processes;

generating automatically one or more biological inferences regarding relationships between chemical or biological molecules and biological processes using results from the Likelihood statistic analysis methods.

22. The method of Claim 21 further comprising a computer readable medium having stored therein instructions for causing a processor to execute the steps of the method.

23. The method of Claim 21 wherein the step of applying Likelihood statistic analysis methods to the connection network includes applying a Likelihood statistic calculated by:

$$L_{AB} = P(A \mid B) * P(\neg A \mid \neg B) * P(B \mid A) * P(\neg B \mid \neg A),$$

wherein A and B are two chemical or biological molecule names which co-occur in one or more database records, wherein $P(A \mid B) \equiv$ (the probability of A given B), $P(B \mid A) \equiv$ (the probability of B given A), wherein $P(\neg A \mid \neg B) \equiv$ (the probability of not A given not B) and $P(\neg B \mid \neg A) \equiv$ (the probability of not B given not A).

24. The method of Claim 21 wherein the chemical or biological molecules and biological processes co-occur in a cell.

25. The method of Claim 21 wherein the plurality of arcs connecting the plurality of nodes in a pre-determined order includes a biological pathway.

26. The method of Claim 21 wherein the step generating automatically one or more biological inferences includes generating a collection of a plurality of chemical or biological molecules logically associated with a plurality of biological process, or a

collection of a plurality of biological processes logically associated with a chemical or biological molecule.

27. The method of Claim 26 wherein the step of generating automatically one or more biological inferences between chemical or biological molecules and a biological process using results from the Likelihood statistic analysis methods includes generating automatically one or more biological inferences between chemical or biological molecules and a biological process in a cell using results from the Likelihood statistic analysis methods.